

	SVM	SVM (RR)	SVM (SDB)	SVM (RM)	LFR [17]
label error	0.1471 (5.7e-17)	0.2007 (0.002)	0.1869 (0.004)	0.1740 (0.003)	0.2299
bias	0.1689 (5.7e-17)	0.0050 (0.003)	0.0036 (0.009)	0.0795 (0.010)	0.0020
RRB	0.2702 (0.014)	0.2926 (0.004)	0.3172 (0.025)	0.2545 (0.007)	n/a
	LR	LR (RR)	LR (SDB)	LR (RM)	DADT [7]
label error	0.1478 (4.8e-04)	0.2077 (0.004)	0.1802 (0.002)	0.1810 (0.003)	0.1600
bias	0.1968 (0.003)	0.0044 (0.006)	0.0060 (0.011)	0.0262 (0.008)	0.0090 (0.015)
RRB	0.4647 (0.013)	0.4696 (0.009)	0.5402 (0.011)	0.4282 (0.019)	n/a
	AdaBoost	AB (RR)	AB (SDB)	AB (RM)	AB (FWL)
label error	0.1529 (0.002)	0.2078 (0.004)	0.1822 (0.005)	0.1864 (0.004)	0.1860 (0.004)
bias	0.1856 (0.012)	0.0091 (0.006)	0.0013 (0.007)	0.0381 (0.013)	0.0682 (0.004)
RRB	0.4372 (0.032)	0.4661 (0.019)	0.5461 (0.015)	0.4410 (0.013)	0.4321 (0.016)

Table 2: A summary of our experimental results for the Census Income data for relabeling, massaging, and the fair weak learner. The threshold for SDB was chosen to achieve perfect statistical parity on the training data. Standard deviations are reported in parentheses when known.

	SVM	SVM (RR)	SVM (SDB)	SVM (RM)	CND [6]
label error	0.2823 (0)	0.2778 (0.025)	0.2979 (0.022)	0.3000 (0.017)	0.2757
bias	0.0886 (4.2e-17)	0.0732 (0.066)	0.0266 (0.085)	0.0445 (0.028)	0.0327
RRB	0.6756 (0.081)	0.7827 (0.054)	0.8619 (0.041)	0.6232 (0.070)	n/a
	LR	LR (RR)	LR (SDB)	LR (RM)	
label error	0.2541 (0.005)	0.2656 (0.020)	0.2685 (0.021)	0.2625 (0.011)	
bias	0.1383 (0.014)	0.0095 (0.064)	0.0142 (0.219)	0.0202 (0.566)	
RRB	0.3070 (0.067)	0.8564 (0.045)	0.8687 (0.042)	0.6741 (0.045)	
	AdaBoost	AB (RR)	AB (SDB)	AB (RM)	AB (FWL)
label error	0.2602 (0.009)	0.2429 (0.010)	0.2745 (0.010)	0.2637 (0.019)	0.2859 (0.016)
bias	0.2617 (0.272)	0.0376 (0.044)	0.0034 (0.064)	0.0391 (0.023)	0.0093 (0.035)
RRB	0.6774 (0.219)	0.8629 (0.051)	0.8596 (0.067)	0.6965 (0.037)	0.6879 (0.042)

Table 3: A summary of our experimental results for the German data for relabeling, massaging, and the fair weak learner. The threshold for SDB was chosen to achieve perfect statistical parity on the training data. On this dataset SVM was run with a linear kernel. Standard deviations are reported in parentheses when known.

	SVM	SVM (RR)	SVM (SDB)	SVM (RM)	
label error	0.2718 (5.7e-17)	0.2793 (0.009)	0.2716 (0.013)	0.2876 (0.015)	
bias	0.0550 (1.4e-17)	0.1460 (0.017)	0.0106 (0.035)	0.0260 (0.047)	
RRB	0.2424 (0.045)	0.2588 (0.009)	0.3064 (0.042)	0.2552 (0.032)	
	LR	LR (RR)	LR (SDB)	LR (RM)	
label error	0.2742 (1.14e-16)	0.3130 (0.011)	0.2745 (0.010)	0.2966 (0.008)	
bias	0.1468 (9.99e-18)	0.3025 (0.040)	0.0034 (0.640)	0.0732 (0.024)	
RRB	0.1971 (0.036)	0.3213 (0.035)	0.8596 (0.067)	0.2117 (0.036)	
	AdaBoost	AB (RR)	AB (SDB)	AB (RM)	AB (FWL)
label error	0.2690 (0.004)	0.3088 (0.009)	0.2990 (0.008)	0.2860 (0.019)	0.2687 (0.008)
bias	0.0966 (0.020)	0.2123 (0.013)	0.0140 (0.017)	0.0180 (0.037)	0.0463 (0.016)
RRB	0.2864 (0.057)	0.3996 (0.105)	0.4027 (0.061)	0.3325 (0.060)	0.2971 (0.028)

Table 4: A summary of our experimental results for the Singles data for relabeling, massaging, and the fair weak learner. The threshold for SDB was chosen to achieve perfect statistical parity on the training data. Standard deviations are reported in parentheses when known.